ETLA Working Papers

No. 10

23 May 2013

Antti Kauhanen
Martti Kulvik
Silja Kulvik
Sirpa Maijanen
Olli Martikainen
Paula Ranta

RESOURCE ALLOCATION IN HEALTH CARE PROCESSES

A case study

Suggested citation: Kauhanen, Antti, Kulvik, Martti, Kulvik, Silja, Maijanen, Sirpa, Martikainen, Olli & Ranta, Paula (23.5.2013). "Resource Allocation in Health Care Processes: A Case Study". ETLA Working Papers No 10. http://pub.etla.fi/ETLA-Working-Papers-10.pdf

ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

ISSN-L 2323–2420
ISSN 2323–2420 (print)
ISSN 2323–2439 (online)

Resource allocation in health care processes: a case study[1]

Antti Kauhanen[a,b], Martti Kulvik[a], Silja Kulvik[a], Sirpa Maijanen[a], Olli Martikainen[a], Paula Ranta[a]

Abstract: This paper utilizes queuing models to analyze health care processes. We extend previous queuing models to allow for i) heterogeneous resources, ii) resource allocation to various tasks, and iii) teams (complementary resources). We model a process of one clinical unit. We use the model to analyze how resource allocation affects both process performance and utilization of resources. This approach emphasizes how allocation of resources to tasks affects process performance. We illustrate how the model can be used to analyze how variations in resources affect process performance and for example how ICT affects process performance.

JEL: I12, C61, D24

Keywords: processes, queues, performance, information and communication technology

Tiivistelmä: Tässä artikkelissa hyödynnetään jonomalleja terveydenhuollon prosessien analysoinnissa. Laajennamme aiempia malleja huomioimalla i) resurssien heterogeenisuuden, ii) resurssien allokoinnin eri tehtäviin, ja iii) tiimit (komplementaariset resurssit). Mallinnamme yhden osaston toiminnan. Analysoimme mallin avulla kuinka resurssien allokointi vaikuttaa sekä prosessin suorituskykyyn että resurssien käyttöasteeseen. Tämä lähestymistapa korostaa prosessien suorituskyvyn riippuvuutta resurssien allokoinnista eri tehtäviin. Näytämme kuinka mallin avulla voidaan analysoida resurssirajoitteiden vaihtelun ja tietotekniikan hyödyntämisen merkitystä prosessien suorituskyvylle.

Avainsanat: prosessit, jonomallit, suorituskyky, informaatio- ja kommunikaatioteknologia

[a] Research Institute of the Finnish Economy

[b] Corresponding author. Address: Lönnrotinkatu 4 B, 00120 Helsinki, Finland. e-mail:

antti.kauhanen@etla.fi

[1] Funding from the Finnish Funding Agency for Technology and Innovation (Tekes) and Technology Industries of Finland Centennial Foundation is gratefully acknowledged.

Introduction

Rising costs of health care have led both researchers and practitioners to pay more attention to efficiency and productivity. To improve health care both reliable performance measures and successful approaches for improvement are needed (Perla et al. 2011) . In a recent study, 265 efficiency measures in health care were compared but almost all of these measures reflected only cost of care, not efficiency (Hussey et al. 2009). The most popular efficiency measures were based on parametric (production function) or non-parametric (DEA or SFA) approaches to finding out the most effective configurations of measured outputs compared to measured inputs (Hollingsworth 2003). These methods consider the system in study as a black box, we know the relation of inputs to outputs, but the mechanisms inside the systems are unknown.

To create improvements to the system in study, information of the mechanisms inside the system and their possible problems is helpful. In manufacturing, there is a long lasting tradition to analyze the organizational capabilities and their fit to the product-process architecture (Fujimoto 2007). In Japanese manufacturing multiskilled labor and organizational problem solving cycles have been essential in developing productivity (Clark and Fujimoto 1991). Traditionally, queuing models have been used in analyzing the industrial production processes (Nyhuis and Wiendahl 2008). Similarly, queuing and operational analysis models can be applied in the performance analysis of organizational processes (Denning and Buzen 1978, Gelenbe and Pujolle 1998).

Measuring, modeling and improving the performance of health care processes is one way to increase their productivity (e.g. Plsek 1997, Locock 2003, Langabeer et al. 2009, de Mast et al. 2011). Analyses of processes may reveal better ways of organizing the work flow, better division of work between employees, redundant tasks, and bottlenecks. Improvements based on such analyses may increase output for fixed resources.

Analysis of health care processes (or workflow) is a demanding task as the processes are complex in nature (Plsek 1997, Malhotra et al. 2007). Consider a clinical unit. There may be several classes of patients, who are in turn treated by groups of clinical practitioners. These groups consist of several different professions. Each clinical practitioner has several tasks during a day, works in multiple groups each day, and the completion of these tasks depend largely on other group members and the actions of other groups. Moreover, the groups work with many patients during a shift. Typically, the clinical units operate 24 hours a day each day of the year.

In health care processes human resources play a large role in determining process performance and payroll form a large share of costs. Quantitative analyses of optimal use of health care resources have often been conducted using queuing analysis (see e.g. Green 2006). These models deal with the problem of determining how many beds, nurses, etc. are needed in a clinical unit for a given flow of patients. Examples of applications are nurse staffing for time varying demand (Green et al. 2007) and the optimal number of beds in an intensive care unit (ICU) (McManus et al. 2004). Most of the models consider only one resource at a time, for example nurses. Yankovic and Green (2011) are an exception, as they study both beds and nurses simultaneously. They note that allowing for heterogeneous resources would be potentially valuable, since in reality two types of nurses are used, and the tasks they are allowed to perform differ.

The previous studies using queuing models have not modeled *tasks* of different clinical practitioners, nor the *skills* needed for various tasks. These considerations are important since nurses for example carry out wide array of tasks, only some of which are related to direct patient care (e.g. Jinks and Hope 2000, Williams et al. 2009). This raises the concern that should nurses or some other employee group carry out these tasks (Jinks and Hope 2000). Many tasks in a clinical unit are such that they require a group of practitioners (e.g. two nurses and a doctor) with specific skills to carry out the task. An important aspect here is that the skills are complementary (for example, the doctor cannot be substituted by a nurse).

The performance of a process depends heavily on allocation of resources to tasks. In other words, who does what is a key determinant of performance. Even though resources are complementary, there are possibilities for substitution. For example, we can replace lower skill resource with a higher skill resource if needed.

Ignoring tasks and the complementarity of resources, the performance of the processes may be overestimated. Complementarities also mean that small changes in resources may lead to large changes in performance. This happens if the bottleneck resource is used in many teams. Conversely large changes in some resources may lead only to small changes in performance if some other resource becomes a bottleneck and is used by many teams.

The analysis of tasks and resources also permits the examination of technological changes on process performance. An example is a technological change that decrease the time spent on some task. The effect of such changes on process performance depends on how the newly freed resources can be used. Again, the complementarity plays a large role in this analysis.

We build on the literature on queuing analyses and health care processes and extend these analyses to allow for i) heterogeneous resources, ii) resource allocation to various tasks, iii) teams (complementary resources). We model a process of one clinical unit. Our model has 2 patient classes; 13 types of clinical practitioners who form 21 teams; more than 90 tasks, some which depend on patient load and others that do not. We also model the allocation of resources to various tasks. An important feature of the model is that it can be used to study how changes in the resource constraints affect process performance. Thus decision makers can use the model to analyze how for example hiring of an additional nurse affects process performance and utilization of resources.

We use the model to analyze how resource allocation affects both process performance and utilization of resources. This approach emphasizes how allocation of resources to tasks affects process performance. We illustrate how the model can be used to analyze how variations in resources affect process performance and for example how ICT affects process performance.

Case description

To build a queuing network model of a clinical ward, we need to characterize the work flow in the ward. In the following we describe the ward, the care process and the work flow. In the next section we describe the mathematical model that is built from the information presented here.

The care process of cerebral vascular disease patients

The case unit is the acute neurology ward 92 of Helsinki University Central Hospital (HUCH). The ward treats mainly stroke patients. The ward consists of two units: The Stroke Unit and the Neurology Ward. The Stroke Unit has five patient beds and it treats acute patients who need constant monitoring. The contiguous 10-bed Neurology Ward treats patients with a more stable condition. A typical patient comes to Stroke Unit from the Emergency Room, stays for a few days, and is then transferred to the Neurology Unit. The patients may, however, be transferred to other hospitals, come directly to the Neurology Ward, or have some other routing as well.

There are 13 different employee categories working in the Ward: 1) nurses, 2) practical nurses, 3) physiotherapists, 4) a speech therapist, 5) a neuropsychologist, 6) an occupational therapist, 7) a social worker, 8) orderlies, 9) a head nurse, 10) a secretary, 11) a pharmacist, 12) doctors, and 13) a senior physician.

The information presented here was gathered by semi-structured interviews. At least one person from each occupational group was interviewed, resulting in a total of 14 interviews. All interviews (average duration was 1.5 hours) were recorded and transcribed. The purpose of the interviews was to collect detailed data of the work tasks of the employees. For the modeling, we need to characterize the tasks, know the duration of the tasks and who can perform different tasks. We also need to know in which order the tasks are performed. In our model we have identified more than 90 tasks.

Patient flow

To get a rough picture of the process, Figure 1 shows a diagram of the patient flow. The patient first enters the Emergency Unit. From there he or she is transferred to the neurological ward 92. Depending on the condition of the patient, she is either allocated to the Stroke Unit or the General Unit. We define three customer classes based on the time they spend at each unit: 1) the thrombolysis patients who stay in the Stroke Unit for only one day (class 1), 2) other patients who need intensive care, but who stay at the Stroke Unit for three days (class 2), and 3) patients who go straight to the general unit without passing by the Stroke Unit (class 3).

The process at the department starts with the arrival process after which the patient stays several days in the system. The following step is the care process, which depends on the class (1, 2, 3) of the patient. The care process ends when the patient is ready to be discharged. When the discharge process ends, the patient leaves the ward.

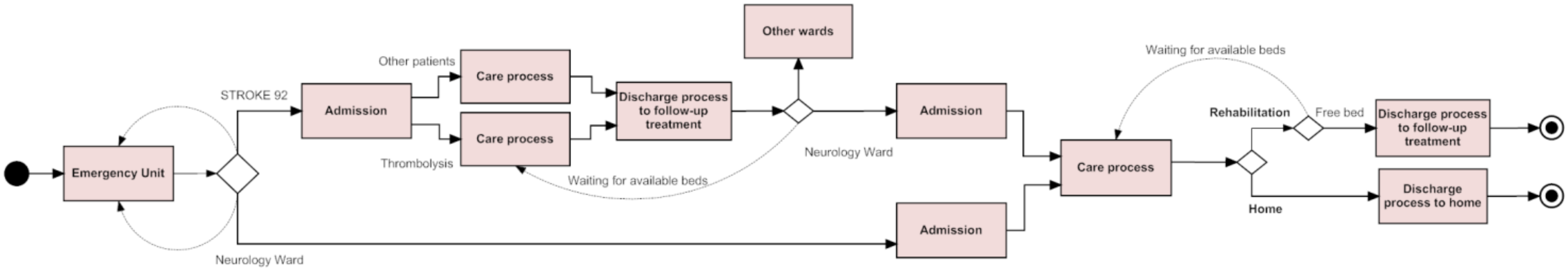

Figure 1

Tasks

Each of the sub-processes (admission, care, discharge) consists of several tasks. The tasks in turn are interconnected because patients move from one task to another according to some probabilities. For each of tasks performed, the key characteristics are the time it takes to perform the tasks and the personnel classes that may carry out the task. These characteristics of the tasks affect fundamentally the possibilities to improve work processes.

The tasks in the sub-processes are only a subset of all tasks carried out in the ward. The tasks in these processes are ones that take place with the patients physically present. The rest of the tasks, including medical orders, regular meetings and other so-called “back office” tasks are not directly related to patient care. However, such tasks are important as a lot of time has to be allocated to such tasks. In effect, these tasks reduce the working time available to other activities.

Human resources and their allocation to tasks

Each of the tasks requires some human resources. Many tasks require a combination of personnel, i.e. teams. For example, medication is always done by the nurse, but for example the morning showers may be performed by one practical nurse or a nurse, or alternatively by two or even three of them, depending on the patient’s condition. In some tasks, the resource requirements of may be fixed, by law or practical needs. In other words, in some tasks resources are substitutable whereas in other tasks they are complements. This is important for process analysis. The nature of tasks and resource requirements has to be taken into account. Otherwise the suggested improvements may violate the resource constraints.

Mathematical formulation[2]

The model

The information presented in the previous section provides sufficient data for defining a stochastic queuing network model of the ward. The key building blocks that are needed are customer arrival intensities (λ_n) to the system for each patient class, patient classes (E_c) served, activities or tasks (A_n), transition probabilities from task n n to task j (p_{nj}), service times in activities (T_n),

[2] This section draws on Naumov and Martikainen (2011a) and Naumov and Martikainen (2011b).

resources (R_l), requirement of resources l in team m (r_{ml}), and service rate of m in activity n (π_{nm}). This information can be gathered by interviews as outlined above.

Based on the process description and the variables above we can define a queuing network model M.

$$M = \left(A_n, E_c, p_{nj}, T_n, \pi_{mn}, r_{ml}\right) \tag{1}$$

The throughput of the network depends on the allocation of teams m to activities n. This allocation is given by the matrix $\mathbf{X}$. The aggregate production of teams allocated to activity n is given by $\eta_n(\mathbf{X}) = \sum_{m=1}^{M} \pi_{nm} x_{nm}$. The network throughput is defined by $\lambda(\mathbf{X}) = \min_{1 \le n \le N} \frac{\eta_n(\mathbf{X})}{w_n}$, where w_n is the total expected workload in task n. Thus, the throughput is determined by the task for which production in relation to workload is the smallest. The total expected workload in turn depends on customer arrival rates and routing probabilities. Thus this variable captures the important aspects of the work flow.

The problem is to choose the matrix $\mathbf{X}$ in order to maximize the throughput of the network while satisfying resource constraints. Formally, the problem is the following:

Maximize

$$\lambda(\mathbf{X}) = \min_{1 \le n \le N} \left(\frac{1}{w_n} \sum_{m=1}^{M} \pi_{nm} x_{nm} \right)$$

subject to given resource constraints:

$$\sum_{n=1}^{N} \sum_{m=1}^{M} x_{nm} r_{ml} \le R_l, \quad l = 1, \ldots, L,$$

$$\sum_{n \in S_i} \sum_{m=1}^{M} x_{nm} \le B_i, \quad i = 1, 2, \ldots, K,$$

$$x_{nm} \ge 0, \quad n = 1, 2, \ldots, N, \quad m = 1, 2, \ldots, M.$$

The first constraint states that we cannot use more of resource l in teams m which serve in task n than is available. The second constraint states that we cannot exceed the maximum number B_i of servers allowed in given subsets S_i of tasks.

Solving the model

The optimal fractional allocations of the teams of resources in the model M with resource constraints is solvable (Naumov and Martikainen 2011b, Naumov and Martikainen 2011a) and gives the optimal solution G which maximizes customer throughputs in tasks λ_{cn}, minimizes the utilizations of resources ρ_l and provides also the utilization of teams ρ_m and the utilization of tasks ρ_n in the system:

$$(\lambda_{cn}, \rho_l, \rho_m, \rho_n) = G(\lambda_n, R_l, M) \qquad (2)$$

The utilization rates measure the fraction of time the resources are employed.

Performance analysis

Solving function G for the model M and the variables λ_n and R_l in Formula (2) does the performance analysis. The calculation reveals for example the optimal allocation of resources to teams that can be assigned to the activities. The joint use of resources that is specified in the teams and the optimization algorithm included enables the analysis of externalities caused by resource sharing. For instance, an improvement in one process releases resources that can be moved to other processes in the organization.

When the processes are analyzed using model M and function G, the modeling results can be calibrated with the process performance data of the real process. The calibration means the comparison of existing real process performance statistics to the corresponding results given by the analysis tools. If the calibration does not succeed, iterative interviews are needed to correct the process diagrams and their variables. This creates more insight of the process behavior. In some cases experimenting with the process variables such as “incorrect delays” or possible “hidden work times” has been needed to reveal and correct the factors that prevent successful calibration. Only after successful calibration the possible process changes can be modeled and their effects analyzed.

Process improvements

Now we can study the impact of different possible changes made in model M . The changes may concern any of the inputs to the model, including the process description. Here we illustrate the analysis of process improvements by considering an abstract alternative model called M_1 .

Before the comparison of the original and improved models can be taken, the functions G have to be solved for each of the models. If resources in teams are used, then also the optimal functions $\underline{G}$ can be obtained for the models. Let us denote the resulting variables of the solution $G(\lambda_n,R_l,M)$ of Formula (2) using the following notation $\lambda_{cn} = G(\lambda_n,R_l,M)(\lambda_{cn})$.

We can analyze several types of improvements. Here we consider three types. The first consider the throughput achieved with models M and M_1 with the same resources. The service level obtained by a customer class c in model M can be expressed as the throughput λ_{cn} for some activity n. The service level or throughput improvement of model M_1 compared to model M is calculated by Formula(3):

$$\Delta\lambda_{cn} = G(\lambda_n,R_l,M_1)(\lambda_{cn}) - G(\lambda_n,R_l,M)(\lambda_{cn}) \tag{3}$$

The second type of improvement is resource savings with constant service level. We obtain the resource improvements $\Delta R_l = \underline{R}_l - R_l$ related to a constant service level λ_{cn} from Formula (4):

$$G(\lambda_n,\underline{R}_l,M_1)(\lambda_{cn}) = G(\lambda_n,R_l,M)(\lambda_{cn}) \tag{4}$$

The third type considers utilization of resources with a constant level of service. The utilization improvement of resources l related to a constant service level λ_n can be calculated from Formula (5):

$$\Delta\rho_l = G(\lambda_n,R_l,M_1)(\rho_l) - G(\lambda_n,R_l,M)(\rho_l) \tag{5}$$

Similar formulas can be written to other variable improvements by keeping some reference variable as constant. The improvements can be expressed as absolute or relative quantities.

Results

Calibrating the model

In addition to the interviews, we received administrative statistics from the Ward 92, which are needed for constructing and calibrating the model. The first model we created did not match the administrative statistics as the constructed process was not able to treat the actual number of patients. The system was severely overloaded. After interviewing the people again, we found out that some of that tasks are done simultaneously. Another remark was that one nurse from the Stroke Unit was sometimes helping in the general unit. After taking these features into account, the model was calibrated and it could care the same amount of patients that the statistics show.

Baseline results

As the model was ran through, the results showed high utilization rates for the resources. The utilization rate measures the workload of the personnel. In practice, utilization rates over 0.7 lead to exhaustion over time. In this model, the rate for the nurses and practical nurses was over 0.9. Also all the other worker classes of which we had sufficient data, were overloaded. Thus the process is operating efficiently, but at its limits. The results suggest that in the case of the smallest unexpected absence or event, the capacity will collapse.

Sensitivity of the process performance to sickness absence

We wanted to examine what happens if one nurse is absent. We reduced the number of nurse resources by one and calculated the model. The result is quite intuitive: the capacity of the process is reduced by 15 per cent. This means that the department is able to process only 85 per cent of the patients compared to the original patient flow. In reality the case is not that straightforward, since there is some flexibility in the quality of the care. When the nurses are in a hurry, the tasks can be done in a shorter time, but they are usually performed less carefully and mistakes are done more easily. Thus the actual capacity decrease is somewhere between 0-15 percent, but together with impaired quality of care.

We also calculated the situation, where two nurses are absent. In this case the capacity is decreased by 60 percent from the original. Only 40 per cent of patients can be treated.

These results suggest that in the case of absences that cannot be covered – which is a real problem during tougher times in hospitals – part of incoming patients have to be turned down.

Patients that would benefit from the specialized care are placed somewhere else in the hospital. This has been reality also in our case department. During a severe flu period in fall 2010, the whole General Unit had to be shut down because there were not enough employees.

It is also worth examining what happens, when one nurse is added to the resources. Interestingly, the capacity is not increased at all. This is due to the fact, that the other resources become bottlenecks. Even if the number of nurses would be enough to treat more patients, the patients need also therapists and doctors. As they also had high utilization rates in the beginning, they become the new bottlenecks. However, the utilization rate of the nurses and practical nurses is decreased close to the optimum. This means that the workers are more satisfied and the quality of care increases. For instance, the nurses have more time to do their tasks with a rehabilitative touch, which is beneficial for the patients' recovery.

Technological change and process performance

As an example of a technological change, we calculated what would happen for process performance if the log-in time to information systems was reduced. In the ward, nurses use on average 30-45 minutes of each shift just to log in to the information systems. In principle, this time could be reduced to zero with for example fingerprint identification.

The results show that utilization rate would be diminished for nurse and especially practical nurses. As above, such change would likely result in increased quality of health care.

Table 1 Maximum arrival rates and utilization rates of different resources

		Utilization rate					
	Max arrival rate	Nurse (stroke)	Practical nurse (stroke)	Nurse	Practical nurse	Doctor	Senior physician
Baseline	0.1	0.934	0.904	0.938	0.961	0.964	0.932
One additional nurse	0.1	0.764	0.597	0.944	0.916	0.964	0.932
One nurse absent	0.085	0.977	0.912	0.977	0.978	0.82	0.792
Two nurses absent	0.04	0.891	0.431	0.495	0.43	0.386	0.373
Reduction in log-in delays	0.1	0.891	0.705	0.932	0.935	0.964	0.932

Discussion and conclusion

In this paper we have applied a new method for joint analysis of process performance and utilization of resources. Service processes are intensive in human resources and quality of service, and quality of work life, depends on the utilization rate of these resources. The method relies on detailed description of processes, tasks, and allocation of resources and team of resources to tasks. These descriptions are not valuable only as input to process analysis, but are useful pieces of information themselves in improving service operations.

The method shows how throughput of the process and utilization of resources varies when the resource constraints are changed. The resource constraints, e.g. the number of nurses available, are a key decision variable of hospital managers. The method can also be used to assess the sensitivity of throughput to changes in the resources. The sensitivity of the process is hard to evaluate without a model, since the resources work in teams.

The method can also be used to study how various process changes, e.g. reduction in times needed carry out some tasks affect throughput and resource utilization. Again, the impacts of the changes are difficult to evaluate without a detailed model, since the utilization of various resources and maximum throughput of the system depend on each other in complex ways.

We have illustrated the method on a single hospital ward. The process functions very close to the maximum performance as determined by resource constraint and structure of the process. Many of the employee groups work under considerable workload. This makes the process very process vulnerable and inflexible. Reductions in key resources quickly bring the whole process to a halt.

The results also show that adding for example nurses does not necessarily help, since other resources quickly become bottle necks. This highlights the need to add resources to right places.

A limitation of this approach is the cost of carrying out the interviews and possible problems with accuracy of interviews. However, in principle almost all of the data needed for the process analysis can be acquired with wireless technology. In practice this means that the personnel will be carrying a smartphone and various places were tasks are carried out have bluetooth transmitters. An application in the smartphone then records the times spent on various tasks. This approach is described in Zhang et al. (2011).

References

Clark, Kim B and Takahiro Fujimoto. 1991. Product Development Performance: Strategy, Organization, and Management in the World Auto Industry: Harvard Business Press.

de Mast, Jeroen, Benjamin Kemper, Ronald J. M. M. Does, Michel Mandjes and Yohan van der Bijl. 2011. "Process Improvement in Healthcare: Overall Resource Efficiency." Quality and Reliability Engineering International, Vol. 27, No. 8, pp. 1095-1106.

Denning, Peter J. and Jeffrey P. Buzen. 1978. "The Operational Analysis of Oueueing Network Models." ACM Computing Surveys, Vol. 10, No. 3, pp. 225-261.

Gelenbe, Erol and Guy Pujolle. 1998. Introduction to Queueing Networks. Hoboken, NJ: Wiley.

Green, Linda. 2006. Queueing Analysis in Healthcare. In Randolph W. Hall. Patient Flow: Reducing Delay in Healthcare Delivery, pp 281-307. New York, NY.: Springer.

Green, Linda V., Peter J. Kolesar and Ward Whitt. 2007. "Coping with Time-Varying Demand When Setting Staffing Requirements for a Service System." Production & Operations Management, Vol. 16, No. 1, pp. 13-39.

Hollingsworth, Bruce. 2003. "Non-Parametric and Parametric Applications Measuring Efficiency in Health Care." Health Care Management Science, Vol. 6, No. 4, pp. 203-218.

Hussey, Peter S, Han De Vries, John Romley, Margaret C Wang, Susan S Chen, Paul G Shekelle and Elizabeth A McGlynn. 2009. "A Systematic Review of Health Care Efficiency Measures." Health services research, Vol. 44, No. 3, pp. 784-805.

Jinks, A. M. and P. Hope. 2000. "What Do Nurses Do? An Observational Survey of the Activities of Nurses on Acute Surgical and Rehabilitation Wards." Journal of Nursing Management, Vol. 8, No. 5, pp. 273-279.

Langabeer, James, Jami DelliFraine, Janelle Heineke and Ibrahim Abbass. 2009. "Implementation of Lean and Six Sigma Quality Initiatives in Hospitals: A Goal Theoretic Perspective." Operations Management Research, Vol. 2, No. 1, pp. 13-27.

Locock, L. 2003. "Healthcare Redesign: Meaning, Origins and Application." Quality and Safety in Health Care, Vol. 12, No. 1, pp. 53-57.

Malhotra, Sameer, Desmond Jordan, Edward Shortliffe and Vimla L. Patel. 2007. "Workflow Modeling in Critical Care: Piecing Together Your Own Puzzle." Journal of Biomedical Informatics, Vol. 40, No. 2, pp. 81-92.

McManus, Michael L., Michael C. Long, Abbot Cooper and Eugene Litvak. 2004. "Queuing Theory Accurately Models the Need for Critical Care Resources." Anesthesiology, Vol. 100, No. 5, pp. 1271-1276.

Naumov, Valeriy and Olli Martikainen. 2011a. "Method for Throughput Maximization of Multiclass Networks with Flexible Servers ". Research Institute of the Finnish Economy, Discussion Papers 1261.

Naumov, Valeriy and Olli Martikainen. 2011b. "Optimal Resource Allocation in Multiclass Networks". Research Institute of the Finnish Economy, Discussion Papers 1262.

Nyhuis, Peter and Hans-Peter Wiendahl. 2008. Fundamentals of Production Logistics: Theory, Tools and Applications: Springer.

Perla, Rocco J, Elizabeth Bradbury and Christina Gunther-Murphy. 2011. "Large-Scale Improvement Initiatives in Healthcare: A Scan of the Literature." Journal for Healthcare Quality, Vol. 35, No. 1, pp. 30-40.

Plsek, Paul E. 1997. "Systematic Design of Healthcare Processes." Quality in Healthcare, Vol. 6, No. 1, pp. 40-48.

Williams, Heather, Ruth Harris and Lynne Turner-Stokes. 2009. "Work Sampling: A Quantitative Analysis of Nursing Activity in a Neuro-Rehabilitation Setting." Journal of Advanced Nursing, Vol. 65, No. 10, pp. 2097-2107.

Yankovic, Natalia and Linda V. Green. 2011. "Identifying Good Nursing Levels: A Queuing Approach." Operations Research, Vol. 59, No. 4, pp. 942-955.

Zhang, Ye, Olli Martikainen, Petri Pulli and Valeriy Naumov (2011). Real-Time Process Data Acquisition with Bluetooth. Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, ACM.